

RAHUL ANAND

MACHINE LEARNING ENGINEER

CONTACT

📞 9834842796
✉ rahulanand1103@gmail.com
🌐 rahulanand.me
🌐 [linkedin/rahul-anand1103](https://www.linkedin.com/in/rahul-anand1103)
📄 [github/rahul-anand1103](https://github.com/rahul-anand1103)

SKILLS

Programming Languages: Python

Web Framework: Flask, Fast-API

Generative AI: Prompt Engineering, VectorDB, Stable-diffusion, GPT, LLAMA, Agent, Langchain, Langgraph, LLM Fine Tuning LLM, OpenAI Realtime API

Machine learning: KNN, Logistic regression, Linear Regression, Naive Bayes, SVM, Decision Tree, Random Forest, Boosting (Xgboost, Catboost & Light Gbm), K-means clustering, PCA
Libraries: Sklearn, Numpy, Pandas, Matplotlib, Seaborn, Dask

NLP: BOW, TF-IDF, Word2vec, Glove, LSTM, Transformer-Based Models

NLP Libraries: NLTK, Gensim, Spacy, Text Blob, Parrot, NlpAug

Deep learning: ANN, NLP, Computer Vision

Deep learning Libraries: TensorFlow, PyTorch

CNN: Le-Net, Alex Net, VGG, ResNet, Transfer Learning, YOLO, RCNN, Faster RCNN, Masked RCNN

MLops: Airflow, Feature Store, MLflow, Docker, DVC

Tools: VS Code, Jupiter Notebook, Google Colaboratory

Other: web scraping

PROFILE

Machine Learning Engineer with 5.5+ years of experience designing and deploying end-to-end ML systems, recommendation engines, and NLP solutions. Expertise in NLP, ML, Generative AI, transformers, and fine-tuning models. Proficient in MLOps tools like Airflow, Docker, and MLflow to build scalable, efficient ML workflows. Passionate about leveraging AI to drive innovation and enhance user experiences.

WORK EXPERIENCE

Machine Learning Engineer at Careervira

Feb 2022-present

- Built core ML infrastructure with Airflow, Feature Store, MLflow, Docker, and DVC for streamlined development and deployment.
- Designed and deployed ML systems for personalized and content-based course recommendations using collaborative filtering and user interaction data, and built course-to-job, job-role-to-course and article-article recommendation systems to enhance personalized learning and career guidance.
- Built ML models to predict skills and topics for partner courses, improving discoverability and recommendations; achieved strong performance with Logistic Regression, SVC (F1: 0.731), RoBERTa (F1: 0.879), and transitioned to LLMs for better accuracy and scalability.
- Implemented a course report pipeline providing insights into courses using LLMs, with PDF generation via Python WeasyPrint.
- Built a search and RAG system enabling queries across courses and job roles, inspired by Perplexity, combining retrieval and generation for enriched information and user experience.
- Currently researching and developing mentor systems with psychometrics to guide students in course selection and career planning.

RA&D(Machine learning) at Teknuance Pvt. Ltd

Dec 2019 - Jan 2022

Product-Sabhae:

- Sabhae is an ensemble of collaboration and communication tools for management processes within an organization.
- Intent Classification: Classified user query intents with 90% accuracy and sub-intents with 93.2% accuracy.
- Custom NER Training: Developed and integrated custom Named Entity Recognition (NER) models using BERT and Bi-LSTM, achieving a 78% F1 score for entity extraction.
- Text Augmentation: Enhanced text data using augmentation techniques like paraphrasing with Parrot and character/word-level augmentation with NlpAug, improving model robustness by 10%.
- Voice Interface: Integrated a voice user interface with the chatbot, enabling hands-free interactions.
- Sentiment Classification: Conducted POC for video meeting summarization and conversation-based sentiment classification, increasing user insights.
- QA and Summarization: Conducted POC for Q&A and summarization of meeting notes using BERT, streamlining information retrieval.


PERSONAL PROJECTS

RAG Citation

 [rahul-anand1103/rag-citation](https://github.com/rahul-anand1103/rag-citation)


RAG Citation is an project that combines Retrieval-Augmented Generation (RAG) with automatic citation generation. This tool is designed to enhance the credibility of RAG-generated content by providing relevant citations for the information used in generating responses.

Youtube script writer

 [rahulanand1103/youtube-script-writer](https://github.com/rahulanand1103/youtube-script-writer)

YouTube Script Writer is an open-source AI agent that generates tailored video scripts based on title, language, and tone. It supports various formats, from short clips to long-form content, streamlining research and writing for creators.

Ytkit

 [rahul-anand1103/ytkit](https://github.com/rahul-anand1103/ytkit)

With ytkit, you can fetch subtitles from entire YouTube channels, playlists, and search results in just a few lines of code! Perfect for RAG, AI applications, and content analysis.

Product-AI studio

- AI-Studio is a scalable data science tools that help user build complex model by dragging and dropping relevant nodes and connection them.
- Researched various machine learning and NLP techniques and algorithms to enhance model performance.
- Pipeline Development: Developed a robust machine learning pipeline for data preprocessing, feature selection, model training and performance measurement.
- Scalable Pipeline: Built a parallel and scalable data science pipeline using Dask, reducing computation time by 30%.

Associate Data Scientist at Techrefic Technologies

Jul 2019 - Dec 2019

- Product - ResumeAI
- Automated the collection of publicly available resumes from various websites, increasing dataset size by 50%.
- **Information Extraction:** Applied NLP and machine learning techniques for extracting key information from resumes, improving parsing accuracy.

Intern at walkover

Jun 2018 - Aug 2018

- Develop API using Flask and MongoDB.
 - Develop cron jobs using Python.
-

EDUCATION

B.E-Computer Science & Engineering

2013-2017

Chandigarh University